International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 2947-2957

https://doi.org/10.70517/ijhsa464245

The Impact of the Integration of Technology and Creativity in Music Education on Future Music Composition

Dongyue Han[1,*]
[1] Music and Drama College, Zhengzhou Sias University, Zhengzhou, Henan, 451100, China
Corresponding authors: (e-mail: hdylw123456@163.com).

Abstract This paper explores the impact of the integration of technology and creativity in music on future music composition, with a focus on the design and application of music generation systems based on deep learning. First, a Markov chain model is constructed to analyse the distribution of musical notes. The CNN-Attention mechanism is then combined to extract the main melody, and an improved Transformer-XL model is used to enhance the quality of music generation. Objective evaluations show that in terms of repetition rate, the improved Transformer-XL model achieved significant optimisation, with a repetition rate of only 17.63%, representing a 48.82% decrease compared to Melody_LSTM. Subjective evaluations revealed that the system achieved an average score of 4.12 across five operational performance dimensions. In terms of music generation quality, the system scored 4.5 and 4.6 on the two key dimensions of style consistency and musical authenticity, respectively, demonstrating a clear advantage over the control system's scores of 3.2 and 3.3.

Index Terms deep learning, music generation, Markov chain model, CNN-Attention, improved Transformer-XL

I. Introduction

Music and technology are intricately linked in many ways, and their mutual promotion and integration continue to drive the advancement of technology and art [1], [2]. With the advent of the digital age, technology has further influenced the art form of music, bringing revolutionary changes to music performance, dissemination, and especially future music creation [3]-[5].

With the continuous advancement of technology, music creators can now utilise more convenient and precise tools to complete their works, thereby creating a broader creative space for creators [6], [7]. This is primarily reflected in the following methods: (1) Digital audio technology: Digital audio technology is a major achievement in the integration of music and technology, enabling the digitisation and electronicisation of music [8], [9]. Through digital audio technology, music can be recorded and edited with greater precision, making the music production process more efficient and flexible [10], [11]. Additionally, digital audio technology has opened up new possibilities for musicians to create more complex and rich musical works [12]. (2) Electronic synthesizers: Electronic synthesizers are another important technological achievement in music, capable of simulating the tones and sound effects of various instruments, playing a significant role in music creation [13], [14]. Through electronic synthesizers, musicians can create various unique sound effects, adding freshness and creativity to musical works [15], [16]. (3) Virtual reality technology: With the continuous development of virtual reality technology, an increasing number of musicians are beginning to apply virtual reality technology to music creation and performance [17], [18]. Through virtual reality technology, musicians can create music works with greater immersion and audiovisual effects, providing audiences with a new musical experience [19], [20]. (4) Technology: Technology is a rapidly developing field in recent years, and it has also brought new possibilities to music creation [21], [22]. Through technology, musicians can use intelligent software for composition and arrangement, achieving automation and intelligence in music creation. This new creative model provides musicians with more inspiration and possibilities [23]-[25].

Reference [26] points out that the application of computer technology has enriched the means and forms of music creation and is also a key factor in the production and development of modern music. Based on this, the advantages of computer technology in music creation are examined, and its application in music creation, arrangement, and other areas is discussed. Literature [27] investigates the impact of technology on music composition, employing ethnographic music survey methods and intuitive interpretation analysis methods to collect data. The results indicate that digital technology has facilitated faster, more convenient, and higher-quality music composition, while also promoting the widespread dissemination of music. Literature [28] explores the relationship between modern music creation and technological development, analysing music and technology in two contexts: music creation and music education. It reveals that human factors constrain the development of music, as people are reluctant to embrace the possibilities offered by new technologies. Literature [29] analyses the design and use of artificial intelligence (AI) music tools for contemporary popular music creation, reporting on the performance of AI tools in music and identifying usage patterns as well as the challenges and issues faced. Literature [30] aims to promote the development of music creation by applying computer technology in music creation to achieve the goal of human-computer interaction. Literature [31] emphasises that the creative application of digital technology has transformed the ways in which music is created, disseminated, and consumed, and has also influenced the quality of music. By analysing creative music production cases, it discusses new usage patterns and personalised music 'consumption' methods. Literature [32] examines the impact of AI technology on music creation and the music industry's economy. By comparing two record companies, it highlights that works with moderate AI involvement have the highest profitability, underscoring the importance of harmonious collaboration between humans and technology. Literature [33] explores the application of AI technology in electronic music and composition. Based on a literature review, it outlines the impact of AI technology on music and analyses the current status of AI technology application in electronic music creation through a questionnaire survey, emphasising that the introduction of intelligent technology in electronic music creation is not given sufficient attention. Literature [34] constructs a folk music creation model based on electronic music technology and tests its functionality, with results validating the feasibility of the aforementioned creation model, which effectively enhances the accuracy of music creation. Literature [35] explores the practical issues encountered in the application of computer music production software in music composition, combining outstanding musical works from China and abroad. The above studies explore the application of AI, digital technology, computer technology, and other scientific and technological advancements in music composition, emphasising that such technologies not only effectively improve the efficiency and quality of music composition but also enhance the economic benefits of musical works.

This paper employs a Markov chain model for song analysis to extract the pitch and duration distribution patterns of musical works. A CNN-Attention network integrating time-domain and channel-domain attention mechanisms is designed to address the issue of feature weight allocation in melody extraction. An improved Transformer-XL model with an enhanced Mask mechanism is proposed to overcome the convergence bottleneck in long-sequence music generation. The effectiveness of the proposed model is validated through objective metrics. The negative log-likelihood method is used to assess the fitting degree of different models. Music students are invited to provide subjective evaluations to examine the performance advantages of the proposed system in terms of operability and music generation quality.

II. Design of a music generation system based on deep learning

Music creation is undergoing a paradigm shift driven by technology. With the rapid development of artificial intelligence technology, the integration of technology and creation in the field of music has become a core issue of concern in academia. To reveal the future development direction of music creation enabled by technology from an interdisciplinary perspective, this paper explores the application paths of deep learning technology in the field of music generation and its artistic value.

II. A. Song analysis based on Markov chains

II. A. 1) Obtaining pitch sequences and duration sequences

In the processed sample melodies, the pitch composition of all notes in each song is extracted to form a pitch sequence. Taking a certain song as an example, the pitch sequence of the monophonic melody obtained is as follows:

$$S_{p_1} = \left\{c^2, a^1, g^1, \cdots, c^1, a\right\} \tag{1}$$

In formula (1), S_{p_1} represents the set of all note pitches in a song. The sequence represents the order and number of pitches.

Using the same method, the note duration sequence of the sample song can be obtained as follows:

$$S_{t_1} = \left\{\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \cdots, \frac{1}{8}, \frac{1}{4}\right\} \tag{2}$$

In formula (2), S_{t_1} represents the set of all note values in the song, where the sequence of numbers indicates the order and quantity of note values, and the note values correspond to the pitch order

II. A. 2) Obtaining note distribution

In this subsection, note distribution includes note pitch and duration. The Markov model needs to obtain the initial distribution of pitch and duration separately.

The initial pitch distribution is obtained under the following conditions:

$$D(p_m) = \frac{pn}{PN} \tag{3}$$

As shown in Formula (3), the initial pitch distribution is the ratio of the number of pitches in a given pitch sequence to the total number of pitches in the sequence. $D(p_m)$ represents the distribution of a given pitch in a pitch sequence, pn represents the number of times a given pitch appears in the pitch sequence, and PN represents the total number of pitches in the pitch sequence.

The initial duration distribution is obtained under the following conditions:

$$D(t_m) = \frac{tn}{TN} \tag{4}$$

As shown in Formula (4), the initial time value distribution is the ratio of the number of occurrences of a specific time value in a time value sequence to the total number of time values in the sequence. $D(t_m)$ represents the distribution of a specific time value in a time value sequence, tn denotes the number of times that time value appears in the sequence, and TN is the total number of time values in the sequence.

After applying the above processing to all sample songs, the initial distribution of notes and time values in the overall data will be obtained, with the total distribution satisfying the following conditions:

$$Dp_k = \frac{\sum pn}{\sum PN} \tag{5}$$

As shown in Formula (5), Dp_k is the distribution of a certain pitch in the pitch sequences of all sample songs. $\sum pn$ is the number of a certain pitch in the pitch sequences of all sample songs, and $\sum PN$ is the total number of pitches in the pitch sequences of the sample songs.

The set SDp is the initial distribution of all note pitches in the entire sample song set.

$$SDp = \{Dp_1, Dp_2, Dp_3, \cdots, Dp_k\} \tag{6}$$

As shown in Formula (6), Dt_k is the distribution of a certain time value in the time value sequences of all sample songs. $\sum tn$ is the number of a certain time value in the time value sequences of all sample songs, and $\sum TN$ is the total number of time values in the time value sequences of the sample songs.

$$Dt_k = \frac{\sum tn}{\sum TN} \tag{7}$$

Set SDt represents the distribution of all note durations in the entire sample song collection.

$$SDt = \{Dt_1, Dt_2, Dt_3, \cdots, Dt_k\} \tag{8}$$

II. B.Main melody extraction based on CNN-Attention

The main melody extraction network in this section is based on an encoder-decoder structure combined with an attention mechanism. Since music data is relatively small, even the spectrograms obtained after data preprocessing are relatively small images. Overly complex network structures can easily lead to overfitting when extracting main melodies. Therefore, the principle of network design in this experiment is to keep it as simple as possible.

The structure of the main melody extraction network is shown in Figure 1. The Encoder-Decoder network of the main melody extraction network consists of three layers for both the encoder and decoder, with each layer comprising an Attention module and an up/down sampling layer. A one-dimensional convolution layer is added at the end to assist in extracting the main melody.

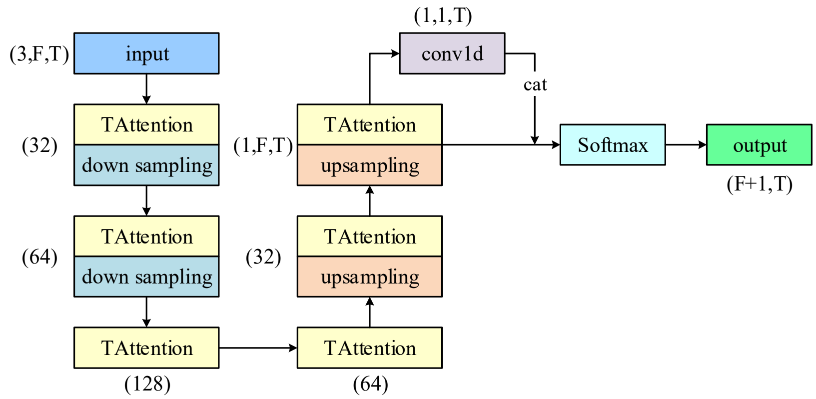

Figure 1: Melody extraction network structure

(1) TAttention module

The TAttention module is shown in Figure 2 and consists of two parts, corresponding to two Attention structures. The attention mechanism has good modelling capabilities for relatively long music data, and since the notes in music are closely related, the attention mechanism can also fully learn the correlations between the notes and distinguish between primary and secondary notes, thereby identifying the notes corresponding to the main melody.

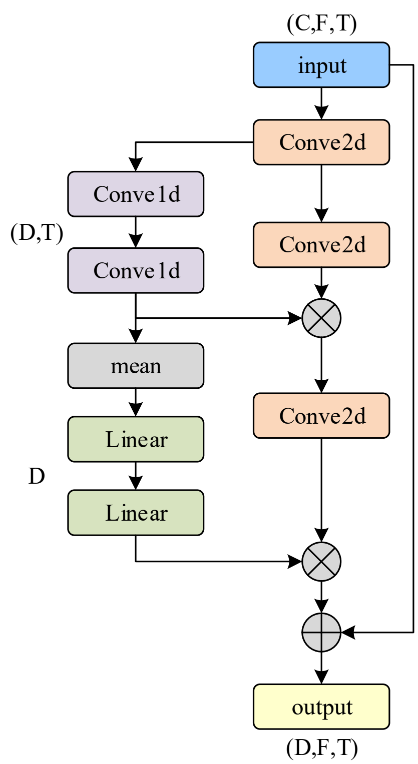

Figure 2: TAttention module

1) Time-domain attention mechanism: In a music sequence, each time point along the time axis either has a main melody or does not. The attention mechanism is used to model the correlations between different time points, and through network learning, the importance of each time point is determined, with different weight coefficients assigned to each time point, thereby emphasising important features and suppressing non-important features. In this section's experiment, we need to identify which time points in the music sequence are important and which are secondary. The introduction of the temporal attention mechanism can effectively improve the accuracy of main melody extraction. As shown in Figure 2, the temporal attention mechanism is implemented using one-dimensional convolution. Specifically, the information is first compressed into the temporal domain, as shown in Formula (9).

This is achieved through a mean operation. Subsequently, one-dimensional convolution is used to learn the weights in the time domain.

$$U = \frac{1}{F}\sum_{i=1}^{F} S_{tj} \tag{9}$$

2) Channel-domain attention mechanism: In a convolutional neural network, each spectral map is initially represented by one or more channels. After passing through different convolution kernels, each channel generates new signals. For example, if each channel of image features uses a 64-kernel convolution, it will produce a matrix with 64 new channels $(F,T,64)$, F,T representing the frequency domain (height) and time domain (width) of the spectral map, respectively. Each channel's feature essentially represents the component of the spectral map under different convolution kernels. Since each signal can be decomposed into components on the kernel function, the contribution of the new 64 channels to key information will vary. If we assign a weight to each channel's signal to represent its relevance to key information, a higher weight indicates greater relevance, meaning that channel deserves more attention. This strategy was also adopted in this section for audio processing. First, different processing methods are applied to the same audio segment to obtain different results, which are then treated as different channels for processing. In this experiment, as shown in Figure 2, the channel-domain attention mechanism is implemented using a fully connected layer, similar to SENet. First, the information is compressed across channels, as shown in Formula (10), which is achieved through a mean operation, and then the weights are learned through the fully connected layer.

$$D = \frac{1}{T}\sum_{i=1}^{T} V_i \tag{10}$$

(2) Tail Convolution

When predicting the main melody, it is necessary to consider the case where the frequency is 0. However, in this section's experiment, the frequency sequence was filtered during data processing, and the filtering was implemented using an exponential function, so the case where the frequency is 0 is missing. Therefore, a tail convolution was specifically added to predict the case where the frequency is 0. The output of the main network is set to $(B,1,F,T)$, and the tail convolution accepts this output. It then compresses F through a one-dimensional convolution, ultimately obtaining an output of $(B,1,1,T)$ representing the case of frequency 0. This is then connected to the output of the main network to produce an output of $(B,1,F+1,T)$ as the final output of the entire main melody extraction network.

(3) Loss function

The loss function of the main melody extraction network proposed in this section is shown in Formula (11).

$$L(y,\hat{y}) = BCE(y,\hat{y}) \tag{11}$$

The binary cross-entropy loss function (BCE) in the loss function is shown in Formula (12).

$$BCE(y,\hat{y}) = -\sum_{i=1}^{N}\left(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right) \tag{12}$$

Among them, $\hat{y}$ represents the network's predicted value, and y represents the true value.

II. C.Music Generation Based on an Improved Transformer-XL Model

Given the advantages of the Transformer-XL model in handling long sequence tasks, this paper selects Transformer-XL as the music generation model and improves its masking mechanism. The Transformer-XL model inherits the traditional Transformer architecture, with an encoder on the left and a decoder on the right. By incorporating a segment-based recurrent mechanism and modifying the relative position encoding within the Transformer structure, it evolves into the Transformer-XL model. This paper further improves the Mask mechanism, and the structure of the improved model is shown in Figure 3.

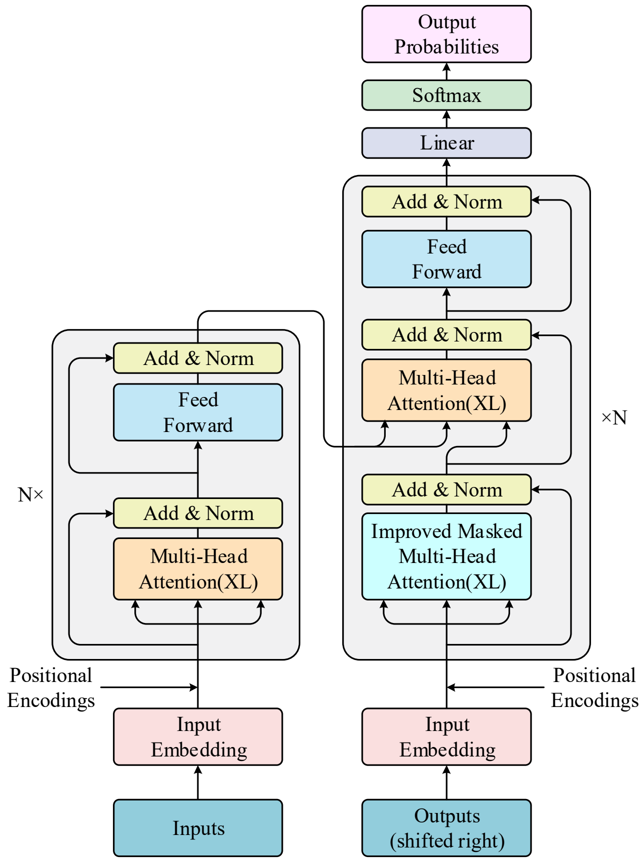

Figure 3: The improved Transformer-XL structure

In the modelling of music, the main framework of the model is based on Transformer-XL, with an improved masking mechanism added to enable the model to see bidirectional information. During training, this paper divides the entire piece of music into multiple segments. The hidden features of the $n-1$ layer of the τ segment of the piece can be represented as:

$$\tilde{h}_{\tau}^{n-1}=\left[SG\left(h_{\tau-1}^{n-1}\right)\circ h_{\tau}^{n-1}\right] \tag{13}$$

Among them, the features of the τ segment utilise the memory features of the previous segment. Although they do not participate in BP calculations, they are also concatenated in dimension, which is equivalent to the current segment reusing the information of the previous segment, increasing the model's dependency, expanding the model's sequence processing length, and effectively solving the fragmentation problem.

When the model is trained, the information of each segment of the input music is represented by the query vector, key vector, and value vector:

$$q_{\tau}^{n},k_{\tau}^{n},v_{\tau}^{n}=h_{\tau}^{n-1}W_{q}^{T},\tilde{h}_{\tau}^{n-1}W_{k}^{T},\tilde{h}_{\tau}^{n-1}W_{v}^{T} \tag{14}$$

In the recursive self-attention calculation, the query vector is represented solely by its own features without the addition of a recursive mechanism, while the key vector and value vector both combine the features of the previous segment. W is the weight that the model needs to learn.

Since the improved Transformer-XL model also uses relative position encoding, the hidden features of the τ layer of the n segment in the model can be calculated using the hidden features of the $n-1$ layer:

$$A_{\tau,i,j}^{n}=q_{\tau,i}^{n\,T}k_{\tau,j}^{n}+q_{\tau,i}^{n\,T}W_{k,R}^{n}R_{i-j}+u^{T}k_{\tau,j}^{n}+v^{T}W_{k,R}^{n}R_{i-j} \tag{15}$$

$$\alpha_{\tau,i}^{n}=\mathrm{Im}\,proved-Mask-Soft\max\left(A_{\tau,i}^{n}\right)v_{\tau,i}^{n} \tag{16}$$

$$\alpha_{\tau}^{n}=\left[\alpha_{\tau,1}^{n}\circ\alpha_{\tau,2}^{n}\circ\cdots\circ\alpha_{\tau,m}^{n}\right]^{T}W^{n} \tag{17}$$

Among them, $\alpha_{\tau,i}^{n}$ is the attention feature of the i head obtained by self-attention calculation of the query vector, key vector, and value vector. After that, it undergoes a series of calculations, including normalisation, residual connection, and positional feedforward network, to obtain the potential feature representation of each position in each segment:

$$z_{\tau}^{n} = LayerNorm\left(Linear\left(\alpha_{\tau}^{n}\right) + h_{\tau}^{n-1}\right) \tag{18}$$

$$h_{\tau}^{n} = \max\left(0, z_{\tau}^{n} W_{1}^{n} + b_{1}^{n}\right) W_{2}^{n} + b_{2}^{n} \tag{19}$$

Among them, h_{τ}^{n} represents the n hidden feature of the τ segment in the piece of music, and the output h_{τ}^{n} can be calculated using the Softmax function to obtain the predicted probability of each note at each moment:

$$P_{\theta} = \frac{\exp\left(h_{\tau,N}^{n}\right)}{\sum_{N} \exp\left(h_{\tau,N}^{n}\right)} \tag{20}$$

Through experimentation, it was found that most music generation models struggle to converge when generating music exceeding 16 bars, and the quality of the generated music deteriorates sharply. Music beyond 16 bars sounds chaotic, as if it were forcibly spliced together. Therefore, many music generation models are unable to generate music with more bars while maintaining quality, resulting in shorter durations. However, this paper improves the mask mechanism in the decoder during the training of the main model, enabling the model to utilise bidirectional information during training and converge when generating music with a larger number of bars. By improving the mask mechanism, the model can generate music with 32 bars or more, extending the duration of the music. After setting the hyperparameters, this paper compared the convergence curves of the improved Transformer-XL model and the Transformer-XL model when generating 32-bar music. By comparing the convergence curves, it was found that after improving the Transformer-XL's Mask mechanism and adding a probabilistic selection method, the model achieves lower training loss and better training performance when generating music with a larger number of bars, thereby demonstrating the effectiveness of the improved Mask mechanism.

III. Application and effectiveness analysis of deep learning-based music generation models

III. A. Experimental results

III. A. 1) Objective evaluation

To validate the effectiveness of the improved Transformer-XL model, this study selected popular music fragments from the Lakh MIDI dataset as sample data. The waveform comparison between the music generated by the improved Transformer-XL model and the sample music is shown in Figure 4. It can be seen that the waveform of the music generated by the improved Transformer-XL model is highly consistent with the sample music in terms of time-domain features, and the model has successfully inherited the style of the sample music.

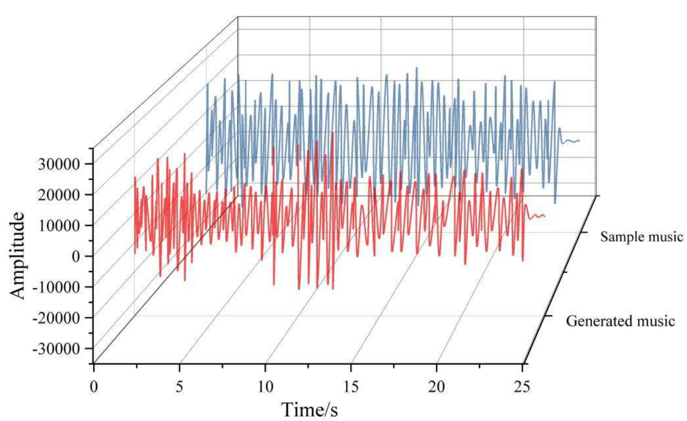

Figure 4: Comparison of Music Waveforms

Pitch distribution can to some extent reflect the similarity in style between two pieces of music, so this experiment requires the statistical analysis of the pitch distribution of each model's output results. The experiment was set up as follows: in each model's 200 compositions, 40 used the same input segment, and among these 20, four models used the same input segment. The final pitch distribution was calculated as the average of the 40 trials. There are two advantages to this approach: first, it effectively controls the influence of irrelevant variables on the experiment; second, it uses multiple trials to eliminate random errors. Both of these advantages make the experimental results more convincing.

MusicalRNN and Original Transformer-XL were used as control models. The pitch distributions of the sample music, the improved Transformer-XL model, and the average pitch distributions of the 40 outputs from the two comparison models are shown in Figure 5. The average pitch distribution of the melodies generated by the improved Transformer-XL model shows the highest similarity to the sample's pitch distribution, followed by the Original Transformer-XL model. This preliminary indicates that the Transformer-XL model has stronger learning capabilities regarding the sample's musical style, and that the improvement strategy proposed in this paper is effective.

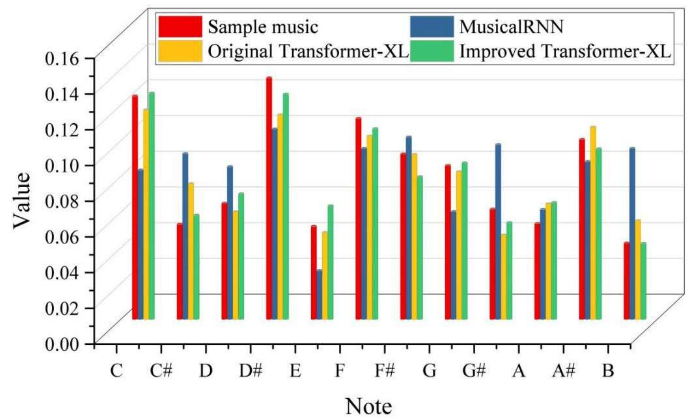

Figure 5: Comparison of average pitch distribution results

In addition, this paper introduces four mainstream music generation models and compares them based on three metrics. The comparison results are shown in Table 1. The three sets of data shown in Table 1 are the average values obtained from 200 trials for each model. The composition model based on Transformer-XL significantly outperforms other models in terms of similarity and repetition rate. Especially in terms of repetition rate, the improved Transformer-XL achieves a substantial optimisation, with only 17.63%, representing a 48.82% decrease compared to Melody_LSTM.

Table 1: Results of the comparison

Model	Similarity	Repetition rate	Yield rate
MusicalRNN	30.18%	30.44%	51%
Original Transformer-XL	36.45%	21.53%	56%
Improved Transformer-XL	39.06%	17.63%	62%
ACMN	28.48%	32.59%	47%
DLGN	26.11%	46.77%	34%
Bidirectional LSTM	16.36%	54.16%	30%
Melody_LSTM	10.17%	66.45%	32%

III. A. 2) Fit degree

The music21 library in Python was used to process the data, and the training cycle was set to 2000 cycles. The change in loss during model training with the training cycle is shown in Figure 6. When the training cycle was 500, the model converged, and the training loss stabilised at around 0.7.

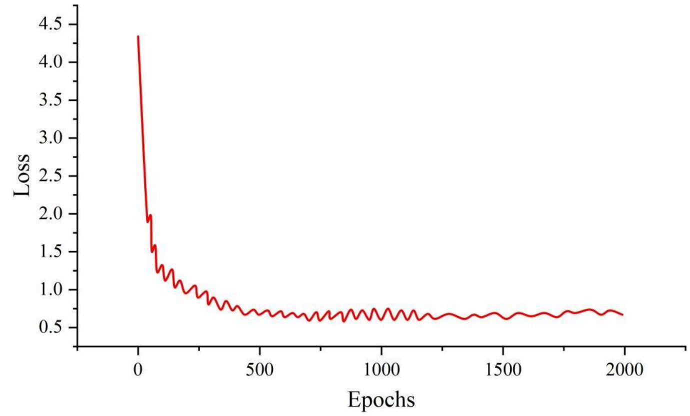

Figure 6: Curve of Loss Variation

MusicalRNN, Original Transformer-XL, and Improved Transformer-XL networks were trained using the same data representation format, and their performance was evaluated. The purpose of maximum likelihood estimation (MLE) is to minimise the cross-entropy between the true data distribution p and the data distribution q generated by the model. By measuring MLE, metrics can be designed to assess the fit between the data and the model. This not only tests the details of the data but also considers the details of the model. Negative Log-Likelihood (NLL) is an improved metric based on Maximum Likelihood Estimation, specifically designed to describe the fit between generated data and real language. The NLL-test loss comparisons for different models are shown in Figure 7. The improved Transformer-XL converges faster and achieves good performance on this metric. The improved Transformer-XL achieves the best NLL performance throughout all stages, while MusicalRNN performs the worst. Due to the similarity of the models, the curves of the original Transformer-XL and the improved Transformer-XL are nearly identical before training round 65. After training round 65, the competitiveness of the original Transformer-XL significantly decreases, further validating the effectiveness of the improvement strategy proposed in this paper. The music generated by the improved Transformer-XL is more closely aligned with real music.

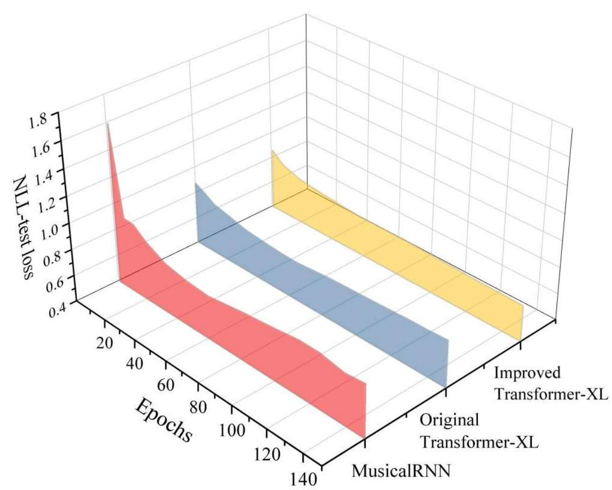

Figure 7: Comparison of NLL-test loss for different models

III. B. Application Effect Analysis

To validate the practical application effectiveness of the music generation system described in this paper, 20 music students were invited to participate in a scoring evaluation. Scoring data was collected using a double-blind testing method, and a 5-point Likert scale (0–5 points) was employed to quantitatively assess the system's various metrics. A mainstream music generation system was selected as the control group to compare the performance of the two systems across five dimensions: user-friendliness, page simplicity, ease of operation, editability, and degree of human involvement (designated as A1–A5, respectively). The comparison results of operational performance in song creation between the two systems are shown in Figure 8. The average score of this system across the five dimensions reached 4.12 points, significantly higher than the control system's 3.08 points.

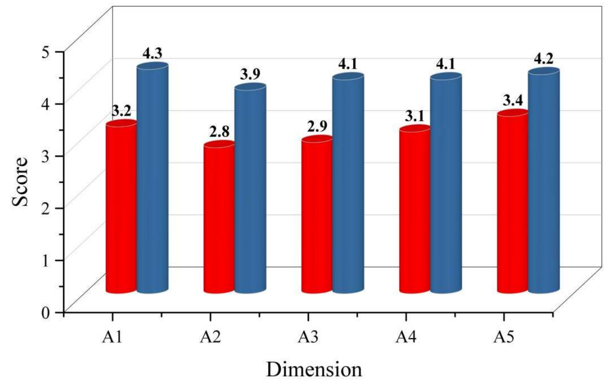

Figure 8: Comparison results of operational performance of different systems

Further evaluation of the performance of the two systems across five dimensions—predefined style quantity, style consistency, detail level, demand matching, and musical authenticity (denoted as B1–B5)—reveals the following results. The comparison of music generation performance between the two systems is shown in Figure 9. Our system achieved scores of 4.5 and 4.6 in the two critical dimensions of style consistency and musical authenticity, respectively, demonstrating a significant advantage over the control system's scores of 3.2 and 3.3. Overall, this confirms the advantages of the improved Transformer-XL model in maintaining musical style consistency and enhancing generation quality, providing reliable technical support for the deep integration of technology and creativity in the field of music.

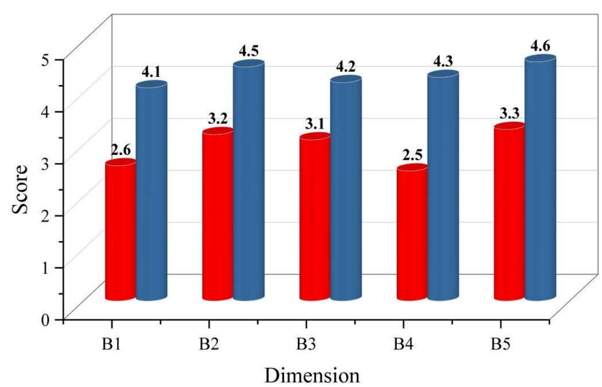

Figure 9: Comparison results of music generation performance of different systems

IV. Conclusion

This study developed a music generation system based on deep learning, demonstrating the feasibility of combining technology with music creation.

The average pitch distribution of the melodies generated by the improved Transformer-XL model shows the highest similarity to the pitch distribution of the sample, followed by the original Transformer-XL. In terms of repetition rate, the improved Transformer-XL has been significantly optimised, with only 17.63%, a decrease of 48.82% compared to Melody_LSTM. The improved Transformer-XL achieved the best NLL scores throughout the entire process, while MusicalRNN performed the worst.

In terms of music creation applications, the average score of this system across the five operational performance dimensions reached 4.12 points, significantly higher than the 3.08 points of the control system. In terms of music generation quality, this system scored 4.5 points and 4.6 points in the two key dimensions of style matching and musical authenticity, respectively, showing a clear advantage over the control system's 3.2 points and 3.3 points.

References

[1] Yoon, J., & Kim, K. (2017). Science song project: Integration of science, technology and music to learn science and process skills. K-12 STEM Education, 3(3), 235-250.

[2] Gorgoretti, B. (2019). The use of technology in music education in North Cyprus according to student music teachers. South African Journal of Education, 39(1).

[3] Born, G. (2022). The dynamics of pluralism in contemporary digital art music. Music and Digital Media: A planetary anthropology, 305-377.

[4] Nijs, L. (2018). Dalcroze meets technology: integrating music, movement and visuals with the Music Paint Machine. Music Education Research, 20(2), 163-183.
[5] Liang, Y. (2025). Collaborative music making in the digital age: fostering creativity in vocal ensembles. Interactive Learning Environments, 33(1), 615-630.
[6] Özdemir, D. (2022). A Conceptual Framework on the Relationship of Digital Technology and Art. International Journal on Social and Education Sciences, 4(1), 121-134.
[7] Kumar, S. (2024). Music and Technology. Sangeet Galaxy, 13(1).
[8] Kladder, J. (2020). Digital audio technology in music teaching and learning: A preliminary investigation. Journal of Music, Technology & Education, 13(2-3), 219-237.
[9] Ashbourn, J. (2020). Audio technology, music, and media. Cham, Switzerland: Springer Nature. doi, 10, 978-3.
[10] Zhang, H., Xu, F., Lyu, K., & Seong, D. (2024). Application of Audio Communication Technology in Music Production and Remote Music Cooperation. International Journal of Communication Networks and Information Security, 16(1), 174-185.
[11] Acil, T. (2024). Re-Thinking Boundaries: The Evolution and Impact of AI in Music and Soundscapes. AVANT. Pismo Awangardy Filozoficzno-Naukowej, (2), 1-15.
[12] Noah, F. K. (2018). The impact of increased accessibility in music technology. Ark Audio, 16.
[13] Yao, S. N. (2019). Audio Effect Units in Mobile Devices for Electric Musical Instruments. IEEE Access, 7, 159239-159250.
[14] Boateng, R., Boateng, S. L., & Budu, J. (Eds.). (2025). AI and the Music Industry: Transforming Production, Platforms, and Practice. CRC Press.
[15] Puentes, P. (2023). The synthesizer in the music education research: A snowball review. Journal of Music, Technology & Education, 16(1-2), 119-131.
[16] De Souza, J. (2024). Modular Synthesizers as Conceptual Models. In Modular Synthesis (pp. 81-105). Focal Press.
[17] Buckley, Z., & Carlson, K. (2019, March). Towards a framework for composition design for music-led virtual reality experiences. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (pp. 1497-1499). IEEE.
[18] Chen, W. (2022). Research on the design of intelligent music teaching system based on virtual reality technology. Computational Intelligence and Neuroscience, 2022(1), 7832306.
[19] Bates, E. (2016). Composing and Producing Spatial Music for Virtual Reality and 360 Media. In Proceedings of the International Festival and Conference on Sound in the Arts, Science and Technology.
[20] Justin, R. S., & Darmayuda, I. K. (2024, December). BOUNDLESS CREATION, VIRTUAL MUSIC AND GLOBAL COLLABORATION. In Proceeding Bali-Bhuwana Waskita: Global Art Creativity Conference (Vol. 4, pp. 55-64).
[21] King, A. (2018). The student prince: music-making with technology. Creativities, technologies, and media in music learning and teaching: an Oxford handbook of music education, 5, 162-178.
[22] Freedman, B. (2017). How Technology Refocuses Music Creation and Composition. The Oxford handbook of technology and music education, 367.
[23] Sturm, B. L., Ben-Tal, O., Monaghan, Ú., Collins, N., Herremans, D., Chew, E., ... & Pachet, F. (2019). Machine learning research that matters for music creation: A case study. Journal of New Music Research, 48(1), 36-55.
[24] Benedict, C., & O'Leary, J. (2019). Reconceptualizing “music making:” Music technology and freedom in the age of neoliberalism. Action, Criticism, and Theory for Music Education, 18(1), 26.
[25] Zhang, M., & Hou, K. (2021, April). Research on the application of computer music making technology in new media environment. In Journal of physics: Conference series (Vol. 1871, No. 1, p. 012142). IOP Publishing.
[26] Liu, C., Wei, L., & Chen, L. (2021, April). Research on the application of computer technology in music creation. In Journal of Physics: Conference Series (Vol. 1883, No. 1, p. 012031). IOP Publishing.
[27] Ikem, G. P. C., & Efurhievwe, M. A. (2022). The functionality of music production technology in the 21st century. Niger Delta Journal of Gender, Peace & Conflict Studies, 2(3), 297.
[28] Cunningham, S., Nicholls, S., & Owens, S. (2017). The development of new technology in creative music applications. In Art, design and technology: Collaboration and implementation (pp. 57-66). Cham: Springer International Publishing.
[29] Deruty, E., Grachten, M., Lattner, S., Nistal, J., & Aouameur, C. (2022). On the development and practice of ai technology for contemporary popular music production. Transactions of the International Society for Music Information Retrieval, 5(1).
[30] Wang, Y., & Zhou, P. (2022, November). Research on the application of computer technology in music creation. In International Conference on Mechanisms and Robotics (ICMAR 2022) (Vol. 12331, pp. 754-759). SPIE.
[31] Danielsen, A. (2017). Music, media and technological creativity in the digital age. Nordic Research in Music Education, 18, 9-22.
[32] Li, S. (2025). The impact of AI-driven music production software on the economics of the music industry. Information Development, 02666669241312170.
[33] Chen, Y. (2021, September). Analysis on the application of artificial intelligence technology in electronic music composition creation. In 2021 4th International Conference on Information Systems and Computer Aided Education (pp. 1200-1204).
[34] Zhang, J. (2025). Research on the Application of Computer-Aided Electronic Music Technology in Folk Music Creation. International Journal of High Speed Electronics and Systems, 34(02), 2440032.
[35] Liang, L., & Liu, J. (2021, April). An exploration of the application of computer music production software in music composition. In 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC) (pp. 794-796). IEEE.